An Algorithm for Unit-Norm Equation Error System Identification Based on the Method of Multipliers

Roberto López-Valcarce, *Member, IEEE*

***Abstract*—It is known that the unit-norm constraint for equation-error based system identification is superior to the monic constraint since it produces consistent estimates for white measurement noise and also presents better approximation properties in reduced-order cases. Here, a new algorithm for unit-norm equation-error adaptive filtering is proposed. This scheme is inspired by the constrained optimization technique known as the method of multipliers. An analysis of stationary points and mean convergence properties is developed.**

***Index Terms*—Adaptive IIR filtering, equation-error, system identification.**

I. Introduction

THE equation-error (EE) approach to system identification offers several advantages over the output error (OE) method in terms of unimodality of the cost function and stability. The EE method estimates the parameters of an unknown system $H(z)$ by adjusting two FIR filters $A(z) = \sum_{i=0}^{M} a_i z^{-i}$ and $B(z) = \sum_{j=0}^{N} b_j z^{-j}$ to minimize the variance of the error signal $e(n)$, as depicted in Fig. 1 for the case $N = M = 3$; the corresponding estimate of $H(z)$ is then given by $B(z)/A(z)$. In spite of those advantages, the traditional monic constraint (in which a_0 is set to one) EE approach has not been widely used since the parameter estimates at convergence are generally biased in the presence of output noise [1]. Early attempts to overcome this problem include instrumental variable techniques [2], for which stability cannot be guaranteed under the same general conditions as for the EE method, and hybrid algorithms [3]–[5], which provide a tradeoff between EE and OE techniques.

Recently, it has been noted that replacing the monic constraint by a unit norm constraint ($\sum_{i=0}^{M} a_i^2 = 1$) in the EE method results in consistent parameter estimates when the measurement noise is white; in addition, the approximation properties of the unit norm solution are superior to those of the monic solution [6]. Consequently, several online adaptive algorithms have been proposed based on this idea. An adaptive unit-norm filter based on planar rotations that, in principle, could be applied to the online problem is described by Regalia in [7]; this parameterization is not convenient, however, if the block $A(z)$ is to be inverted into $A^{-1}(z)$. Davila [8] develops an RLS-like online algorithm, whereas different approaches for developing gradient-based algorithms were presented in [9]–[12].

Manuscript received August 27, 1999; revised February 3, 2003. The associate editor coordinating the review of this paper and approving it for publication was Prof. Scott C. Douglas.

The author was with the Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242 USA. He is now with the Department of Signal Theory and Communications, University of Vigo, Vigo 36200, Spain (e-mail: valcarce@gts.tsc.uvigo.es).

Digital Object Identifier 10.1109/TSP.2003.819012

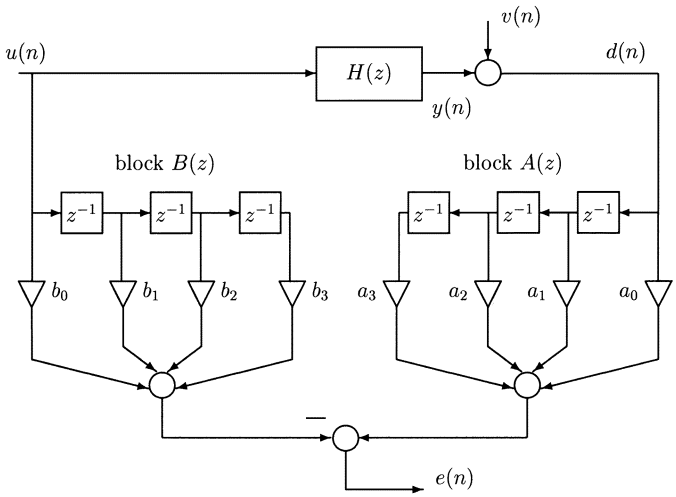

Fig. 1. Equation error configuration ($N = M = 3$).

We consider an alternative technique in order to develop an online algorithm for the unit-norm constraint. Our approach is inspired by the constrained optimization scheme known as the method of multipliers [13] and results in a new LMS-like algorithm. The main idea is to include an additional adaptive parameter that plays the role of a Lagrange multiplier and to adapt simultaneously all the filter coefficients $b_0, \ldots, b_N, a_0, \ldots, a_M$ freely without imposing any constraints. This produces a division-free algorithm since normalization steps are not required, in contrast with several existing methods [10]–[12]. Another attractive feature of our approach is that the concept can be applied to any adaptive filtering problem in which the filter coefficients are required to meet some constraint upon convergence.

The paper is organized as follows. The method of multipliers is briefly reviewed in Section II, and then, the new algorithm is presented. Its stationary points and convergence properties are analyzed in Section III. Simulations are presented in Section IV.

II. New Algorithm

Consider the EE system identification structure depicted in Fig. 1. Here, $u(\cdot)$ represents the input to the unknown system, and $v(\cdot)$ is an additive noise process independent of $u(\cdot)$, which contaminates $y(\cdot)$, the system output, to form the available signal $d(\cdot)$. Both $u(\cdot)$ and $v(\cdot)$ are assumed to be zero-mean wide-sense stationary. The equation error is then $e(n) = A(z)d(n) - B(z)u(n)$, and $A(z), B(z)$ are adapted in order to minimize $E[e^2(n)]$. In order to avoid the trivial

solution $A(z) = B(z) = 0$, some constraints need to be placed. The traditional approach was to fix $a_0 = 1$; more adequate is the unit norm constraint $\sum_{i=0}^{M} a_i^2 = 1$. The equation error is $e(n) = \mathbf{d}_n^t\mathbf{a} - \mathbf{u}_n^t\mathbf{b}$, where

$$\begin{aligned}\mathbf{a} &= [\,a_0 \quad a_1 \quad \ldots \quad a_M\,]^t \\ \mathbf{b} &= [\,b_0 \quad b_1 \quad \ldots \quad b_N\,]^t \\ \mathbf{d}_n &= [\,d(n) \quad d(n-1) \quad \ldots \quad d(n-M)\,]^t \\ \mathbf{y}_n &= [\,y(n) \quad y(n-1) \quad \ldots \quad y(n-M)\,]^t \\ \mathbf{u}_n &= [\,u(n) \quad u(n-1) \quad \ldots \quad u(n-N)\,]^t.\end{aligned}$$

Using $\mathbf{R}_{dd} = E[\mathbf{d}_n\mathbf{d}_n^t]$, $\mathbf{R}_{uu} = E[\mathbf{u}_n\mathbf{u}_n^t]$, and $\mathbf{R}_{ud} = E[\mathbf{u}_n\mathbf{d}_n^t]$, the cost $J(\mathbf{a},\mathbf{b}) = E[e^2(n)]$ is

$$J(\mathbf{a},\mathbf{b}) = \mathbf{a}^t\mathbf{R}_{dd}\mathbf{a} + \mathbf{b}^t\mathbf{R}_{uu}\mathbf{b} - 2\mathbf{a}^t\mathbf{R}_{ud}^t\mathbf{b}. \tag{1}$$

If $v(\cdot)$ is white with variance σ_v^2, then (1) takes the form

$$J(\mathbf{a},\mathbf{b}) = \left(\mathbf{a}^t\mathbf{R}_{yy}\mathbf{a} + \mathbf{b}^t\mathbf{R}_{uu}\mathbf{b} - 2\mathbf{a}^t\mathbf{R}_{uy}^t\mathbf{b}\right) + \sigma_v^2\mathbf{a}^t\mathbf{a} \tag{2}$$

where $R_{yy} = E[\mathbf{y}_n\mathbf{y}_n^t]$ and $\mathbf{R}_{uy} = E[\mathbf{u}_n\mathbf{y}_n^t]$. It is clear from (2) that if the constraint $\mathbf{a}^t\mathbf{a} = 1$ is imposed, then the minimum of J does not change with σ_v^2. Thus, we would like to solve

$$\text{minimize} \quad J(\mathbf{a},\mathbf{b}) = E[e^2(n)] \qquad \text{subject to} \quad \mathbf{a}^t\mathbf{a} = 1.$$

Note that since no constraint is placed on $\mathbf{b}$, the optimum $\mathbf{b}_*$ is immediately found to be

$$\mathbf{b}_* = \mathbf{R}_{uu}^{-1}\mathbf{R}_{ud}\mathbf{a}. \tag{3}$$

[The matrix $\mathbf{R}_{uu}$ is nonsingular provided that $u(\cdot)$ is persistently exciting of sufficient order.] Substituting this back in (1), one obtains a *reduced* cost function in terms of $\mathbf{a}$: $J(\mathbf{a},\mathbf{b}_*) = \mathbf{a}^t\mathbf{R}\mathbf{a}$, where $\mathbf{R} = \mathbf{R}_{dd} - \mathbf{R}_{ud}^t\mathbf{R}_{uu}^{-1}\mathbf{R}_{ud}$. Therefore, the problem becomes

$$\text{minimize} \quad E[e^2(n)] = \frac{\mathbf{a}^t\mathbf{R}\mathbf{a}}{\mathbf{a}^t\mathbf{a}} \tag{4}$$

which is the Rayleigh quotient of $\mathbf{R}$. The minimum value of the cost is the smallest eigenvalue of $\mathbf{R}$, $\lambda_{\min}(\mathbf{R})$, and the optimum vector $\mathbf{a}_*$ is the (unit norm) corresponding eigenvector [6].

Consider now the general constrained optimization problem

$$\text{minimize} \quad f(\mathbf{x}) \qquad \text{subject to} \quad h(\mathbf{x}) = \mathbf{0} \tag{5}$$

with $f : \mathcal{R}^n \to \mathcal{R}$ and $h : \mathcal{R}^n \to \mathcal{R}^m$. The *augmented Lagrangian* is defined as

$$L_c(\mathbf{x},\lambda) = f(\mathbf{x}) + \lambda^t h(\mathbf{x}) + \frac{c}{2}\|h(\mathbf{x})\|^2$$

where $c > 0$, and λ is an m-dimensional vector (the *Lagrange multiplier*). The method of multipliers constructs two sequences of vectors $\{\mathbf{x}_k\}$ and $\{\lambda_k\}$ as follows [13]:

$$\begin{aligned}\mathbf{x}_{k+1} &= \text{argmin}_{\mathbf{x}} L_{c_k}(\mathbf{x},\lambda_k) \\ \lambda_{k+1} &= \lambda_k + c_k \cdot h(\mathbf{x}_{k+1})\end{aligned}$$

where $\{c_k\}$ is a positive penalty parameter sequence. Thus, at the kth step, an unconstrained minimization problem parameterized by the Lagrange multiplier estimate λ_k is solved; then, λ_k is updated, and the iteration is repeated. If $\mathbf{x}_*$ is a solution of (5) and λ_* is the associated Lagrange multiplier, it is one's hope that $\mathbf{x}_k \to \mathbf{x}_*$ and $\lambda_k \to \lambda_*$ as $k \to \infty$. Often (but not always), this requires $c_k \to \infty$ [13].

In the context of the unit norm EE problem, $\mathbf{x} = [\,\mathbf{a}^t \quad \mathbf{b}^t\,]^t$, $f(\mathbf{x}) = J(\mathbf{a},\mathbf{b})$, and $h(\mathbf{x}) = 1 - \mathbf{a}^t\mathbf{a}$. Thus, the augmented Lagrangian takes the form

$$L_c(\mathbf{a},\mathbf{b},\lambda) = J(\mathbf{a},\mathbf{b}) + \lambda(1 - \mathbf{a}^t\mathbf{a}) + \frac{c}{2}(1 - \mathbf{a}^t\mathbf{a})^2. \tag{6}$$

The unconstrained minimization step of the method of multipliers can be performed by means of stochastic gradient descent. From (6), one has

$$\nabla_{\mathbf{b}} L_c = \nabla_{\mathbf{b}} J \tag{7}$$

$$\nabla_{\mathbf{a}} L_c = \nabla_{\mathbf{a}} J - 2\lambda\mathbf{a} - 2c(1 - \mathbf{a}^t\mathbf{a})\mathbf{a}. \tag{8}$$

For online adaptation, the terms $\nabla_{\mathbf{a}} J$, $\nabla_{\mathbf{b}} J$ can be approximated, as usual, by $2e(n)\mathbf{d}_n$ and $2e(n)\mathbf{u}_n$, respectively. Similarly, the adaptation of the Lagrange multiplier can be performed at every iteration, that is, with each new data pair $\{u(n), d(n)\}$. The resulting algorithm is

$$\begin{aligned}\mathbf{a}_{n+1} &= \mathbf{a}_n - \mu[e(n)\mathbf{d}_n - \lambda_n\mathbf{a}_n - c(1 - \mathbf{a}_n^t\mathbf{a}_n)\mathbf{a}_n] \\ \mathbf{b}_{n+1} &= \mathbf{b}_n + \mu e(n)\mathbf{u}_n \\ \lambda_{n+1} &= \lambda_n + c(1 - \mathbf{a}_{n+1}^t\mathbf{a}_{n+1})\end{aligned}$$

where $\mu > 0$ is the stepsize. We observe that in this adaptive context the parameter c plays the role of a stepsize for λ_n. The algorithm can be simplified by assuming that μ is small so that we can alter the sequence of updates; in addition, the choice $c = \mu/2$ will prove useful for analysis purposes. With these modifications, one obtains

$$\lambda_{n+1} = \lambda_n + \frac{\mu}{2}(1 - \mathbf{a}_n^t\mathbf{a}_n) \tag{9}$$

$$\mathbf{a}_{n+1} = (1 + \mu\lambda_{n+1})\mathbf{a}_n - \mu e(n)\mathbf{d}_n \tag{10}$$

$$\mathbf{b}_{n+1} = \mathbf{b}_n + \mu e(n)\mathbf{u}_n. \tag{11}$$

The vector $\mathbf{b}_n$ is adapted via the standard LMS algorithm, whereas the adaptation of $\mathbf{a}_n$ can be viewed as an LMS update with leakage, where the leakage coefficient is dynamically updated as well. An appealing feature of algorithm (9)–(11) is that it does not use any divisions; in the case that an estimate $\hat{y}(n)$ of the system output $y(n)$ is needed, then only one extra division per iteration is required:

$$\hat{y}(n) = \frac{1}{a_0(n)}\left(\sum_{j=0}^{M} b_j(n)u(n-j) - \sum_{i=1}^{M} a_i(n)\hat{y}(n-i)\right). \tag{12}$$

III. Algorithm Analysis

A. Stationary Points

The stationary points (SPs) of (9)–(11) are the values $(\mathbf{a}_*, \mathbf{b}_*, \lambda_*)$ for which the expectations of the update terms in (9)–(11) vanish. Since $e(n) = \mathbf{d}_n^t\mathbf{a} - \mathbf{u}_n^t\mathbf{b}$, any SP must satisfy

$$1 - \mathbf{a}_*^t\mathbf{a}_* = 0 \tag{13}$$

$$\lambda_*\mathbf{a}_* - \mathbf{R}_{dd}\mathbf{a}_* + \mathbf{R}_{ud}^t\mathbf{b}_* = \mathbf{0} \tag{14}$$

$$\mathbf{R}_{ud}\mathbf{a}_* - \mathbf{R}_{uu}\mathbf{b}_* = \mathbf{0}. \tag{15}$$

From (13), at any SP $\mathbf{a}_*$ must have unit-norm, as expected. Comparing (15) and (3), we also see that at any SP, $\mathbf{b}_*$ is optimized as a function of $\mathbf{a}_*$. Now, isolating $\mathbf{b}_*$ in (15) and substituting this into (14), we have

$$\lambda_*\mathbf{a}_* - \mathbf{R}_{dd}\mathbf{a}_* + \mathbf{R}_{ud}^t\mathbf{R}_{uu}^{-1}\mathbf{R}_{ud}\mathbf{a}_* = \mathbf{0}$$

that is, $\mathbf{R}\mathbf{a}_* = \lambda_*\mathbf{a}_*$, from which we conclude that the SPs are the eigenpairs $(\lambda_*, \mathbf{a}_*)$ of the matrix $\mathbf{R}$, with $\mathbf{a}_*$ having unit norm and $\mathbf{b}_*$ optimized in terms of $\mathbf{a}_*$.

B. Stability Analysis

We proceed now to study the stability properties of the SP corresponding to $\lambda_* = \lambda_{\min}(\mathbf{R})$ by using the ordinary differential equation (ODE) method [14]. Let $\theta_n = [\mathbf{a}_n^t \quad \mathbf{b}_n^t \quad \lambda_n]^t$. Then, under the small stepsize assumption, we can approximate $\lambda_{n+1} \approx \lambda_n$ in (10) so that the algorithm can be compactly written as $\theta_{n+1} = \theta_n + \mu F(\theta_n, u, d)$ with

$$F(\theta_n, u, d) = \begin{pmatrix} \lambda_n\mathbf{a}_n - \mathbf{d}_n e(n) \\ \mathbf{u}_n e(n) \\ \frac{1}{2}(1 - \mathbf{a}_n^t\mathbf{a}_n) \end{pmatrix}.$$

Under general conditions, it can be shown that θ_n converges to the solution of the following ODE in some probabilistic sense [14]:

$$\dot{\theta}(t) = E[F(\theta(t), u, d)] \tag{16}$$

where $t = n\mu$. Now, let $\theta_* = [\mathbf{a}_*^t \quad \mathbf{b}_*^t \quad \lambda_*]^t$ be an SP, i.e., $E[F(\theta_*, u, d)] = \mathbf{0}$. The ODE (16) can be linearized in a neighborhood of θ_* (thus yielding a local analysis):

$$\dot{\theta}(t) = \nabla E[F(\theta_*, u, d)]^t(\theta(t) - \theta_*).$$

The SP θ_* is locally stable if and only if all the eigenvalues of the matrix $\mathbf{B}(\theta_*) = \nabla E[F(\theta_*, u, d)]^t$ lie in the left-hand complex semiplane. Now, differentiating

$$E[F(\theta, u, d)] = \begin{pmatrix} \lambda\mathbf{a} - \mathbf{R}_{dd}\mathbf{a} + \mathbf{R}_{ud}^t\mathbf{b} \\ \mathbf{R}_{ud}\mathbf{a} - \mathbf{R}_{uu}\mathbf{b} \\ \frac{1}{2}(1 - \mathbf{a}^t\mathbf{a}) \end{pmatrix}$$

with respect to θ, one obtains

$$\mathbf{B}(\theta) = \nabla E[F(\theta, u, d)]^t = \begin{bmatrix} \lambda\mathbf{I} - \mathbf{R}_{dd} & \mathbf{R}_{ud}^t & \mathbf{a} \\ \mathbf{R}_{ud} & -\mathbf{R}_{uu} & \mathbf{0} \\ -\mathbf{a}^t & \mathbf{0}^t & 0 \end{bmatrix}.$$

If we let $\tilde{\theta} = \theta_* - \theta = [\tilde{\mathbf{a}}^t \quad \tilde{\mathbf{b}}^t \quad \tilde{\lambda}]^t$ be the parameter deviation vector, then the linearized ODE can be rewritten as $\mathrm{d}\tilde{\theta}(t)/\mathrm{d}t = \mathbf{B}(\theta_*)\tilde{\theta}(t)$. Then, one has the following result.

Theorem 1: The stationary point θ_* of the algorithm (9)–(11) corresponding to $\lambda_* = \lambda_{\min}(\mathbf{R})$ is locally stable if $\lambda_{\min}(\mathbf{R})$ is simple.

Proof: First, observe that $\mathbf{B}(\theta_*)$ is negative semidefinite (even if $\lambda_{\min}(\mathbf{R})$ is multiple):

$$\begin{aligned} &\tilde{\theta}^t\mathbf{B}(\theta_*)\tilde{\theta} \qquad (17)\\ &= \begin{bmatrix} \tilde{\mathbf{a}} \\ \tilde{\mathbf{b}} \\ \tilde{\lambda} \end{bmatrix}^t \begin{bmatrix} \lambda_*\mathbf{I} - \mathbf{R}_{dd} & \mathbf{R}_{ud}^t & \mathbf{a}_* \\ \mathbf{R}_{ud} & -\mathbf{R}_{uu} & \mathbf{0} \\ -\mathbf{a}_*^t & \mathbf{0}^t & 0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{a}} \\ \tilde{\mathbf{b}} \\ \tilde{\lambda} \end{bmatrix} \\ &= \lambda_*\tilde{\mathbf{a}}^t\tilde{\mathbf{a}} - E[(\mathbf{d}_n^t\tilde{\mathbf{a}})^2] + 2E[(\mathbf{d}_n^t\tilde{\mathbf{a}})(\mathbf{u}_n^t\tilde{\mathbf{b}})] - E[(\mathbf{u}_n^t\tilde{\mathbf{b}})^2] \\ &= \lambda_*\tilde{\mathbf{a}}^t\tilde{\mathbf{a}} - J(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}). \qquad (18) \end{aligned}$$

In view of (4), $J(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) \geq \tilde{\mathbf{a}}^t\mathbf{R}\tilde{\mathbf{a}}$. Therefore

$$\begin{aligned} \tilde{\theta}^t\mathbf{B}(\theta_*)\tilde{\theta} &\leq \lambda_*\tilde{\mathbf{a}}^t\tilde{\mathbf{a}} - \tilde{\mathbf{a}}^t\mathbf{R}\tilde{\mathbf{a}} \\ &= -\tilde{\mathbf{a}}^t(\mathbf{R} - \lambda_*\mathbf{I})\tilde{\mathbf{a}} \qquad (19) \end{aligned}$$

which is nonpositive since $\lambda_* = \lambda_{\min}(\mathbf{R})$ yields $\mathbf{R} - \lambda_*\mathbf{I} \geq \mathbf{0}$. Moreover, for $\tilde{\theta} \neq \mathbf{0}$, we can have $\tilde{\theta}^t\mathbf{B}(\theta_*)\tilde{\theta} = 0$ only in the following cases.

1) If $\tilde{\mathbf{a}} \neq \mathbf{0}$, then in view of (19), $\tilde{\theta}^t\mathbf{B}(\theta_*)\tilde{\theta} = 0$ iff $\tilde{\mathbf{a}}$ is an eigenvector of $\mathbf{R}$ associated with $\lambda_{\min}(\mathbf{R})$.
2) If $\tilde{\mathbf{a}} = \mathbf{0}$, then by (18), $\tilde{\theta}^t\mathbf{B}(\theta_*)\tilde{\theta} = -J(\mathbf{0}, \tilde{\mathbf{b}}) = -E[(\mathbf{u}_n^t\tilde{\mathbf{b}})^2]$, which is zero iff $\tilde{\mathbf{b}} = \mathbf{0}$. In that case, $\tilde{\theta}$ must be of the form $\tilde{\theta} = \lambda_{\min}(\mathbf{R})\mathbf{e}$, with $\mathbf{e} = [0 \quad \cdots \quad 0 \quad 1]^t$.

Consider now the Lyapunov function $V(\tilde{\theta}) = (1/2)\tilde{\theta}^t\tilde{\theta}$, whose time derivative is $\dot{V}(\tilde{\theta}(t)) = \tilde{\theta}^t\mathbf{B}(\theta_*)\tilde{\theta} \leq 0$. This shows that for a given initial perturbation $\tilde{\theta}(t_0)$ (small enough so that the local linearization remains valid), the norm of the perturbation $\tilde{\theta}$ cannot increase with time.

We cannot yet conclude that θ_* is a stable SP since it could happen that for some τ, $\tilde{\theta}(\tau) \neq \mathbf{0}$ is such that $\dot{V}(\tilde{\theta}(\tau)) = \tilde{\theta}^t\mathbf{B}(\theta_*)\tilde{\theta} = 0$. This would mean that $V(\tilde{\theta})$ as a function of t passes through a maximum, a minimum, or an inflection point at $t = \tau$. The maximum and minimum can be ruled out since $\dot{V} \leq 0$ for all t. For the inflection points, there are two possibilities: Either $V(t)$ resumes decreasing for $t > \tau$, as desired, or $V(t)$ remains constant (and thus, $\dot{V}(t) = 0$) for $t > \tau$. We show next that this cannot happen if $\lambda_{\min}(\mathbf{R})$ is simple.

If $V(\tilde{\theta}) = 1/2\|\tilde{\theta}\|^2$ remains constant, it means either that $\tilde{\theta}$ has converged to a point (which therefore must be an SP of the ODE) or that $\tilde{\theta}(t)$ wanders in a sphere centered at the origin. At the same time, we know that either these points are of the form $\tilde{\theta} = \tilde{\lambda}\mathbf{e}$, or they satisfy $\mathbf{R}\tilde{\mathbf{a}} = \lambda_{\min}(\mathbf{R})\tilde{\mathbf{a}}$.

Suppose $\tilde{\theta} = \tilde{\lambda}\mathbf{e}$ with $\tilde{\lambda} \neq 0$; then, $\theta = [\mathbf{a}_* \quad \mathbf{b}_* \quad (\lambda_* - \tilde{\lambda})]^t$. If this is to be an SP of the ODE, then $(\lambda_* - \tilde{\lambda}, \mathbf{a}_*)$ must be an eigenpair of $\mathbf{R}$. This is not possible since $\mathbf{a}_*$ cannot be an eigenvector associated to two different eigenvalues. In addition, observe that the intersection of the manifold $\tilde{\theta} = \tilde{\lambda}\mathbf{e}$ with a sphere centered at the origin consists only of two isolated points.

On the other hand, suppose that $\mathbf{R}\tilde{\mathbf{a}} = \lambda_{\min}(\mathbf{R})\tilde{\mathbf{a}}$. Since by definition $\mathbf{R}\mathbf{a}_* = \lambda_{\min}(\mathbf{R})\mathbf{a}_*$, it follows that $\mathbf{R}\mathbf{a} = \lambda_{\min}(\mathbf{R})\mathbf{a}$, i.e., $(\lambda_{\min}(\mathbf{R}), \mathbf{a})$ is an eigenpair of $\mathbf{R}$. This can yield an SP if $\|\mathbf{a}\| = 1$ (for example, if $\mathbf{a} = -\mathbf{a}_*$). However, this is as good as the original SP. If $\lambda_{\min}(\mathbf{R})$ is not simple, $\mathbf{a}$ may wander in a manifold of eigenvectors with unit magnitude (without affecting the performance of the algorithm). That is not possible if $\lambda_{\min}(\mathbf{R})$ is simple: Note that any vector θ_0 whose $\mathbf{a}$-component is $-\mathbf{a}_*$ (the only unit-norm eigenvector associated with $\lambda_{\min}(\mathbf{R})$ different from $\mathbf{a}_*$) satisfies $\|\theta_* - \theta_0\| \geq \|\mathbf{a}_* - (-\mathbf{a}_*)\| = 2\|\mathbf{a}_*\| = 2$. Therefore, for a sufficiently small perturbation, the trajectory of θ cannot pass through such a θ_0.

To sum up, we have that $\dot{V}(\tilde{\theta}(t)) \leq 0$ and that $\dot{V}(\tilde{\theta}(t)) = 0$ can happen only at isolated time instants. Therefore, $V(\tilde{\theta}(t)) \to 0$, implying that $\tilde{\theta} \to \mathbf{0}$; the SP corresponding to the smallest eigenvalue is locally stable. ■

Theorem 1 ensures local stability of the SP of interest when $\lambda_{\min}(\mathbf{R})$ is simple. When $\lambda_{\min}(\mathbf{R})$ is multiple, it follows from

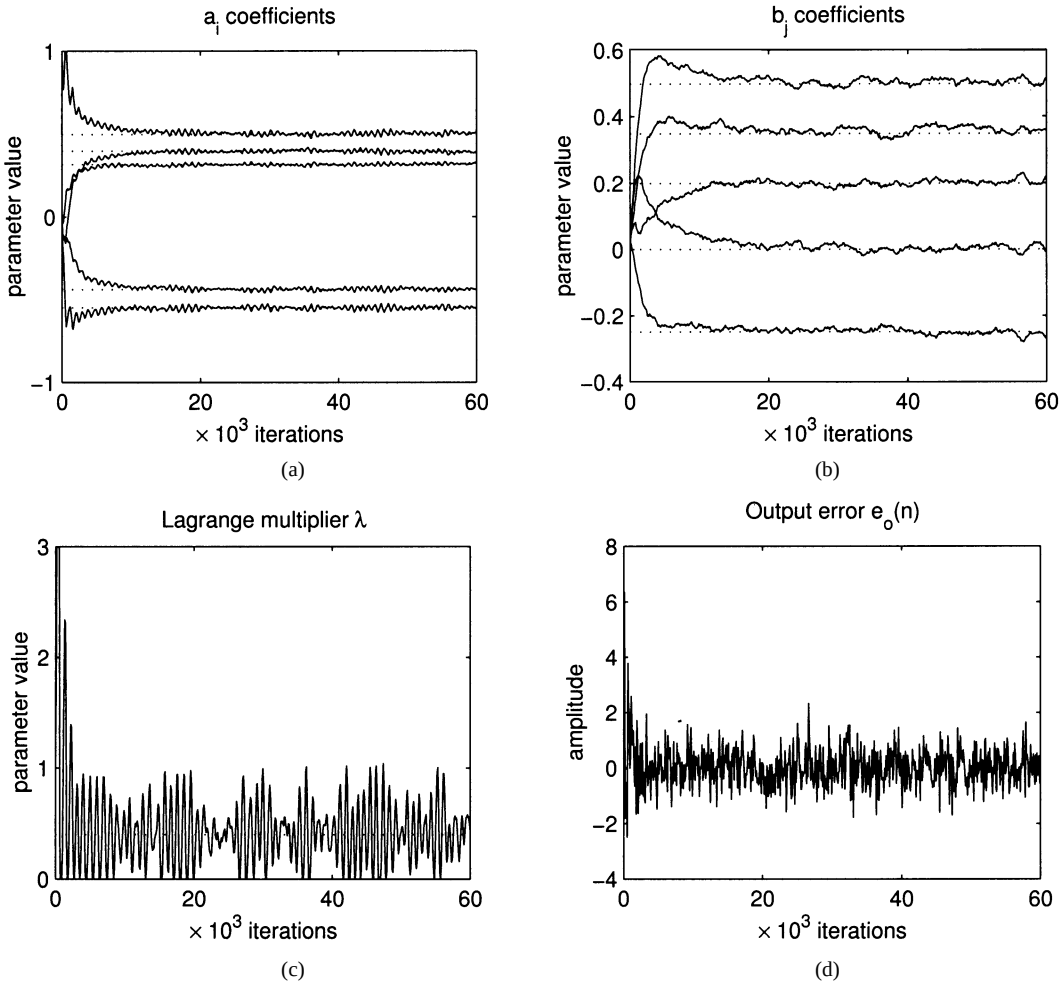

Fig. 2. Results of a single run of the new algorithm. Dashed lines indicate the true parameter values.

the proof above that the parameter estimate may wander in a manifold of eigenvectors; nevertheless, since all eigenvectors provide the same cost reduction, this behavior is acceptable. Finally, Theorem 1 does not provide information about the SPs corresponding to the eigenvalues other than $\lambda_{\min}(\mathbf{R})$. For the maximum eigenvalue, the same reasoning that led to (19) shows that the corresponding matrix $\mathbf{B}(\theta_*)$ is positive semidefinite so that the corresponding SP is never stable. With respect to the remaining eigenvalues, simulation results suggest that the corresponding SPs are not stable either, although a formal proof is not available at this time.

IV. Simulations

A system identification experiment was carried out in order to verify the general performance of the algorithm. The fourth-order system given by

$$H(z) = \frac{1 + 0.4z^{-2} - 0.5z^{-3} + 0.7z^{-4}}{1 - 1.1z^{-1} + 0.8z^{-2} - 0.88z^{-3} + 0.64z^{-4}} \quad (20)$$

is to be identified. The input signal $u(\cdot)$ is a zero-mean white process with unit variance.

The value of the stepsize for the parameters a_i and b_j was $\mu = 5 \times 10^{-4}$. Although the analysis of Section III-B, which showed local stability of the stationary point, is valid only if the stepsize for the Lagrange multiplier is half of that of the coefficients, i.e., $c = \mu/2$, it was found experimentally that the performance of the algorithm is best when c is chosen a few orders of magnitude greater than μ; in this experiment, we set $c = 5 \times 10^{-2}$. This has the effect of speeding up parameter convergence for a fixed μ.

Another feature that proved useful in the practical implementation of the algorithm was to restrict the evolution of the multiplier to the region $\lambda \geq 0$. This is because the value of λ upon convergence is just the value of the cost function J at the stationary point, which is always non-negative. Thus, it makes sense to avoid excursions of λ into the negative region.

The output noise is a white process with variance $\sigma_v^2 = 0.4$, giving a signal-to-noise ratio of 10.8 dB in the reference $d(\cdot)$. All the adaptive coefficients are initially set to zero, except for $a_0(0) = 1$. Fig. 2 shows the results obtained in a single run of the algorithm (9)–(11). As expected, the adaptive coefficients converge in the mean to their true values despite the presence of output noise. In addition, the parameter λ_n is seen to fluctuate around its theoretical value $\lambda_* = J(\mathbf{a}_*, \mathbf{b}_*) = \sigma_v^2$. This fluctuation translates to the oscillations of the adaptive filter coefficients seen in the figure. In order to assess the influence of this behavior in the overall misadjustment, the algorithm was compared with other unit-norm adaptive schemes, namely, those of Ho and Chan [10], Douglas and Rupp [12], and Gao *et al.*

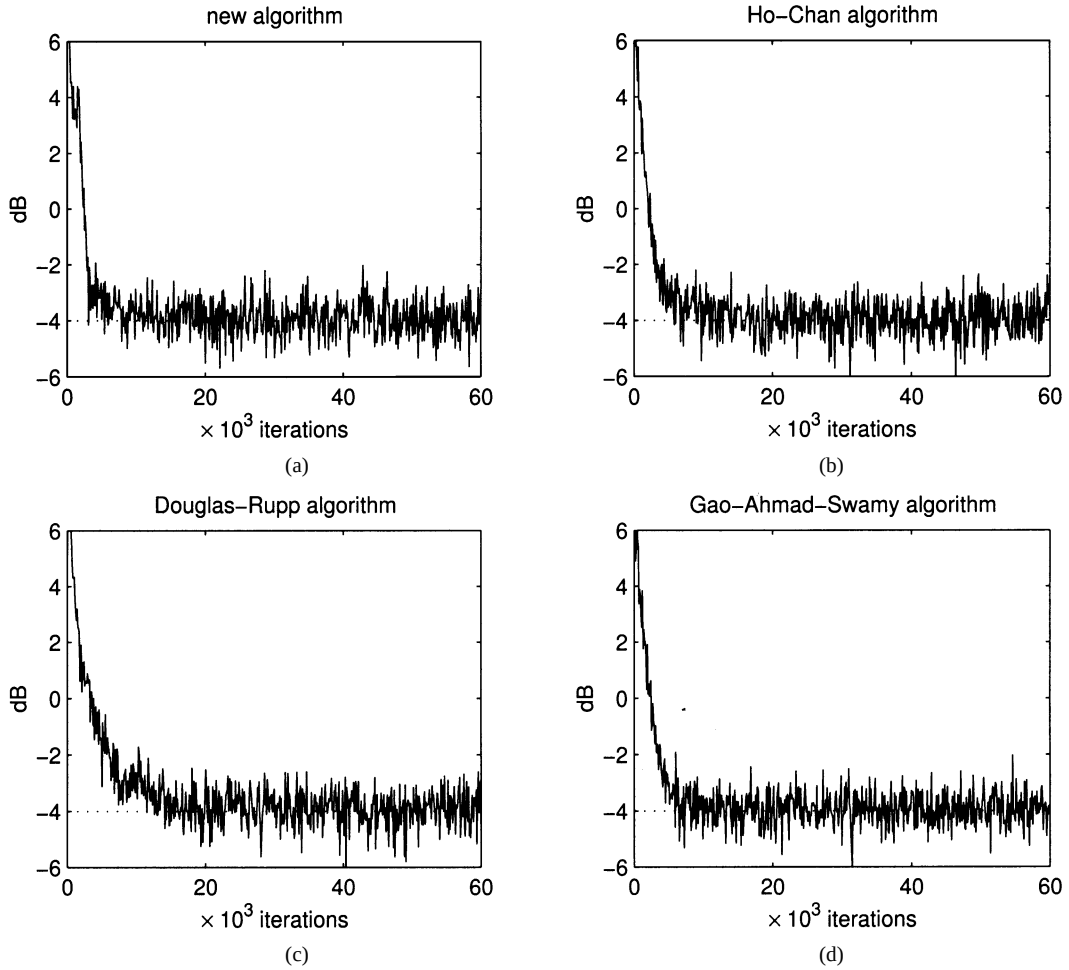

Fig. 3. Output error learning curves SNR $= 10.8$ dB. Dashed lines indicate the noise floor.

[9]. All the algorithms used a value of $\mu = 5 \times 10^{-4}$. Fig. 3 shows the learning curves of the four algorithms, which reflect the mean squared value of the output error $e_o(n) = d(n) - \hat{y}(n)$ averaged over 100 independent runs. It is observed that the oscillation phenomenon in the proposed algorithm does not seem to induce any noticeable increase in misadjustment compared with the other schemes.

Fig. 4 shows the evolution of the squared norm of the parameter error vector, averaged over 100 trials. In terms of this measure, the new scheme is seen to perform slightly better in this example than the algorithms of Douglas and Rupp and Gao *et al.*. The best performance is obtained by the computationally more costly algorithm of Ho and Chan, which was implemented with explicit normalization at every iteration.

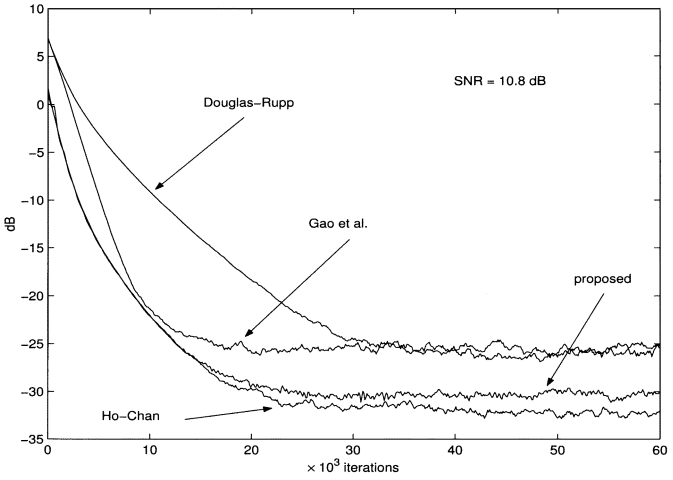

Fig. 4. Comparison of the parameter error vector norm evolution.

V. Conclusions

A new adaptive algorithm for equation-error system identification has been presented. This algorithm is based on the idea of imposing a unit norm constraint on the filter coefficients in order to obtain consistent estimates. Its main novelty is the inclusion of a new adaptive parameter, which plays the role of a Lagrange multiplier. An appealing feature is that there is no need for normalization steps, and therefore, divisions are avoided. The stationary points of the method were analyzed, concluding that the point of interest is locally stable. The idea behind this technique can be applied to any adaptive filtering problem where the filter coefficients are required to meet some constraints upon convergence. Further study should address the theoretical implications of using $c \neq \mu/2$ as well as a mean-squared parameter convergence analysis.

REFERENCES

[1] J. J. Shynk, "Adaptive IIR filtering," *IEEE Signal Processing Mag.*, vol. 6, pp. 4–21, Apr. 1989.

[2] T. Söderström and P. Stoica, *Instrumental Variable Methods for System Identification*, Germany: Springer-Verlag, 1983.

[3] J. B. Kenney and C. E. Rohrs, "The composite regressor algorithm for IIR adaptive systems," *IEEE Trans. Signal Processing*, vol. 41, pp. 617–628, Feb. 1993.

[4] J.-N. Lin and R. Unbehauen, "Bias-remedy least-mean-square equation error algorithm for IIR parameter recursive estimation," *IEEE Trans. Signal Processing*, vol. 40, pp. 62–69, Jan. 1992.

[5] J. E. Cousseau and P. S. R. Diniz, "A general consistent equation-error algorithm for adaptive IIR filtering," *Signal Process.*, vol. 56, pp. 121–134, 1997.

[6] P. A. Regalia, "An unbiased equation error identifier and reduced order approximations," *IEEE Trans. Signal Processing*, vol. 42, pp. 1397–1412, June 1994.

[7] ——, "An adaptive unit norm filter with applications to signal analysis and Karhunen-Loève transformations," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 646–649, May 1990.

[8] C. E. Davila, "An algorithm for efficient, unbiased, equation-error IIR adaptive filtering," *IEEE Trans. Signal Processing*, vol. 42, pp. 1221–1226, May 1994.

[9] K. Gao, M. O. Ahmad, and M. N. S. Swamy, "A constrained anti-Hebbian learning algorithm for total least-squares estimation with applications to adaptive FIR and IIR filtering," *IEEE Trans. Circuits Syst. II*, vol. 41, pp. 718–729, Nov. 1994.

[10] K. C. Ho and Y. T. Chan, "Bias removal in equation-error adaptive IIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 51–62, Jan. 1995.

[11] H.-N. Kim and W.-J. Song, "Unbiased equation-error adaptive IIR filtering based on monic normalization," *IEEE Signal Processing Letters*, vol. 6, pp. 35–37, Feb. 1999.

[12] S. C. Douglas and M. Rupp, "On bias removal and unit-norm constraints in equation-error adaptive IIR filters," in *Proc. 30th Asilomar Conf. Signals Syst., Comput.*, Pacific Grove, CA, Nov. 1996, pp. 1093–1097.

[13] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

[14] H. Fan, "Application of Benveniste's convergence results in the study of adaptive IIR filtering algorithms," *IEEE Trans. Inform. Theory*, vol. 34, pp. 692–709, July 1988.

Roberto López-Valcarce (M'01) was born in Spain in 1971. He received the telecommunications engineer degree from Universidad de Vigo, Vigo, Spain in 1995, and the M.S. and Ph.D. degrees in electrical engineering from the University of Iowa, Iowa City, in 1998 and 2000 respectively.

From 1995 to 1996, he was a systems engineer with Intelsis. He was the recipient of a Fundación Pedro Barrié de la Maza fellowship for continuation of studies. He is currently a research associate (Ramón y Cajal fellow) with the Signal Theory and Communications Department at Universidad de Vigo. His research interests are in adaptive signal processing, communications, and traffic monitoring systems.